Chapter 3

Random Variables and Probability Distributions

3.1 Concept of a Random Variable

Statistics is concerned with making inferences about populations and population characteristics. Experiments are conducted with results that are subject to chance. The testing of a number of electronic components is an example of a **statistical experiment**, a term that is used to describe any process by which several chance observations are generated. It is often important to allocate a numerical description to the outcome. For example, the sample space giving a detailed description of each possible outcome when three electronic components are tested may be written

$$S = \{NNN, NND, NDN, DNN, NDD, DND, DDN, DDD\},$$

where N denotes nondefective and D denotes defective. One is naturally concerned with the number of defectives that occur. Thus, each point in the sample space will be *assigned a numerical value* of 0, 1, 2, or 3. These values are, of course, random quantities *determined by the outcome of the experiment*. They may be viewed as values assumed by the *random variable X*, the number of defective items when three electronic components are tested.

Definition 3.1: | A **random variable** is a function that associates a real number with each element in the sample space.

We shall use a capital letter, say X, to denote a random variable and its corresponding small letter, x in this case, for one of its values. In the electronic component testing illustration above, we notice that the random variable X assumes the value 2 for all elements in the subset

$$E = \{DDN, DND, NDD\}$$

of the sample space S. That is, each possible value of X represents an event that is a subset of the sample space for the given experiment.

Example 3.1: Two balls are drawn in succession without replacement from an urn containing 4 red balls and 3 black balls. The possible outcomes and the values y of the random variable Y, where Y is the number of red balls, are

Sample Space	y
RR	2
RB	1
BR	1
BB	0

Example 3.2: A stockroom clerk returns three safety helmets at random to three steel mill employees who had previously checked them. If Smith, Jones, and Brown, in that order, receive one of the three hats, list the sample points for the possible orders of returning the helmets, and find the value m of the random variable M that represents the number of correct matches.

Solution: If S, J, and B stand for Smith's, Jones's, and Brown's helmets, respectively, then the possible arrangements in which the helmets may be returned and the number of correct matches are

Sample Space	m
SJB	3
SBJ	1
BJS	1
JSB	1
JBS	0
BSJ	0

In each of the two preceding examples, the sample space contains a finite number of elements. On the other hand, when a die is thrown until a 5 occurs, we obtain a sample space with an unending sequence of elements,

$$S = \{F, NF, NNF, NNNF, \dots\},$$

where F and N represent, respectively, the occurrence and nonoccurrence of a 5. But even in this experiment, the number of elements can be equated to the number of whole numbers so that there is a first element, a second element, a third element, and so on, and in this sense can be counted.

There are cases where the random variable is categorical in nature. Variables, often called *dummy* variables, are used. A good illustration is the case in which the random variable is binary in nature, as shown in the following example.

Example 3.3: Consider the simple condition in which components are arriving from the production line and they are stipulated to be defective or not defective. Define the random variable X by

$$X = \begin{cases} 1, & \text{if the component is defective,} \\ 0, & \text{if the component is not defective.} \end{cases}$$

Clearly the assignment of 1 or 0 is arbitrary though quite convenient. This will become clear in later chapters. The random variable for which 0 and 1 are chosen to describe the two possible values is called a **Bernoulli random variable**.

Further illustrations of random variables are revealed in the following examples.

Example 3.4: Statisticians use **sampling plans** to either accept or reject batches or lots of material. Suppose one of these sampling plans involves sampling independently 10 items from a lot of 100 items in which 12 are defective.

Let X be the random variable defined as the number of items found defective in the sample of 10. In this case, the random variable takes on the values $0, 1, 2, \ldots, 9, 10$.

Example 3.5: Suppose a sampling plan involves sampling items from a process until a defective is observed. The evaluation of the process will depend on how many consecutive items are observed. In that regard, let X be a random variable defined by the number of items observed before a defective is found. With N a nondefective and D a defective, sample spaces are $S = \{D\}$ given $X = 1$, $S = \{ND\}$ given $X = 2$, $S = \{NND\}$ given $X = 3$, and so on.

Example 3.6: Interest centers around the proportion of people who respond to a certain mail order solicitation. Let X be that proportion. X is a random variable that takes on all values x for which $0 \leq x \leq 1$.

Example 3.7: Let X be the random variable defined by the waiting time, in hours, between successive speeders spotted by a radar unit. The random variable X takes on all values x for which $x \geq 0$.

Definition 3.2: If a sample space contains a finite number of possibilities or an unending sequence with as many elements as there are whole numbers, it is called a **discrete sample space**.

The outcomes of some statistical experiments may be neither finite nor countable. Such is the case, for example, when one conducts an investigation measuring the distances that a certain make of automobile will travel over a prescribed test course on 5 liters of gasoline. Assuming distance to be a variable measured to any degree of accuracy, then clearly we have an infinite number of possible distances in the sample space that cannot be equated to the number of whole numbers. Or, if one were to record the length of time for a chemical reaction to take place, once again the possible time intervals making up our sample space would be infinite in number and uncountable. We see now that all sample spaces need not be discrete.

Definition 3.3: If a sample space contains an infinite number of possibilities equal to the number of points on a line segment, it is called a **continuous sample space**.

A random variable is called a **discrete random variable** if its set of possible outcomes is countable. The random variables in Examples 3.1 to 3.5 are discrete random variables. But a random variable whose set of possible values is an entire interval of numbers is not discrete. When a random variable can take on values

on a continuous scale, it is called a **continuous random variable**. Often the possible values of a continuous random variable are precisely the same values that are contained in the continuous sample space. Obviously, the random variables described in Examples 3.6 and 3.7 are continuous random variables.

In most practical problems, continuous random variables represent *measured* data, such as all possible heights, weights, temperatures, distance, or life periods, whereas discrete random variables represent *count* data, such as the number of defectives in a sample of k items or the number of highway fatalities per year in a given state. Note that the random variables Y and M of Examples 3.1 and 3.2 both represent count data, Y the number of red balls and M the number of correct hat matches.

3.2   Discrete Probability Distributions

A discrete random variable assumes each of its values with a certain probability. In the case of tossing a coin three times, the variable X, representing the number of heads, assumes the value 2 with probability 3/8, since 3 of the 8 equally likely sample points result in two heads and one tail. If one assumes equal weights for the simple events in Example 3.2, the probability that no employee gets back the right helmet, that is, the probability that M assumes the value 0, is 1/3. The possible values m of M and their probabilities are

m	0	1	3
$P(M = m)$	$\frac{1}{3}$	$\frac{1}{2}$	$\frac{1}{6}$

Note that the values of m exhaust all possible cases and hence the probabilities add to 1.

Frequently, it is convenient to represent all the probabilities of a random variable X by a formula. Such a formula would necessarily be a function of the numerical values x that we shall denote by $f(x)$, $g(x)$, $r(x)$, and so forth. Therefore, we write $f(x) = P(X = x)$; that is, $f(3) = P(X = 3)$. The set of ordered pairs $(x, f(x))$ is called the **probability function**, **probability mass function**, or **probability distribution** of the discrete random variable X.

Definition 3.4: | The set of ordered pairs $(x, f(x))$ is a **probability function**, **probability mass function**, or **probability distribution** of the discrete random variable X if, for each possible outcome x,

1. $f(x) \geq 0$,

2. $\sum_x f(x) = 1$,

3. $P(X = x) = f(x)$.

Example 3.8: A shipment of 20 similar laptop computers to a retail outlet contains 3 that are defective. If a school makes a random purchase of 2 of these computers, find the probability distribution for the number of defectives.

Solution: Let X be a random variable whose values x are the possible numbers of defective computers purchased by the school. Then x can only take the numbers 0, 1, and

2. Now

$$f(0) = P(X = 0) = \frac{\binom{3}{0}\binom{17}{2}}{\binom{20}{2}} = \frac{68}{95}, \quad f(1) = P(X = 1) = \frac{\binom{3}{1}\binom{17}{1}}{\binom{20}{2}} = \frac{51}{190},$$

$$f(2) = P(X = 2) = \frac{\binom{3}{2}\binom{17}{0}}{\binom{20}{2}} = \frac{3}{190}.$$

Thus, the probability distribution of X is

x	0	1	2
$f(x)$	$\frac{68}{95}$	$\frac{51}{190}$	$\frac{3}{190}$

Example 3.9: If a car agency sells 50% of its inventory of a certain foreign car equipped with side airbags, find a formula for the probability distribution of the number of cars with side airbags among the next 4 cars sold by the agency.

Solution: Since the probability of selling an automobile with side airbags is 0.5, the $2^4 = 16$ points in the sample space are equally likely to occur. Therefore, the denominator for all probabilities, and also for our function, is 16. To obtain the number of ways of selling 3 cars with side airbags, we need to consider the number of ways of partitioning 4 outcomes into two cells, with 3 cars with side airbags assigned to one cell and the model without side airbags assigned to the other. This can be done in $\binom{4}{3} = 4$ ways. In general, the event of selling x models with side airbags and $4 - x$ models without side airbags can occur in $\binom{4}{x}$ ways, where x can be 0, 1, 2, 3, or 4. Thus, the probability distribution $f(x) = P(X = x)$ is

$$f(x) = \frac{1}{16}\binom{4}{x}, \quad \text{for } x = 0, 1, 2, 3, 4.$$

There are many problems where we may wish to compute the probability that the observed value of a random variable X will be less than or equal to some real number x. Writing $F(x) = P(X \leq x)$ for every real number x, we define $F(x)$ to be the **cumulative distribution function** of the random variable X.

Definition 3.5: The **cumulative distribution function** $F(x)$ of a discrete random variable X with probability distribution $f(x)$ is

$$F(x) = P(X \leq x) = \sum_{t \leq x} f(t), \quad \text{for } -\infty < x < \infty.$$

For the random variable M, the number of correct matches in Example 3.2, we have

$$F(2) = P(M \leq 2) = f(0) + f(1) = \frac{1}{3} + \frac{1}{2} = \frac{5}{6}.$$

The cumulative distribution function of M is

$$F(m) = \begin{cases} 0, & \text{for } m < 0, \\ \frac{1}{3}, & \text{for } 0 \leq m < 1, \\ \frac{5}{6}, & \text{for } 1 \leq m < 3, \\ 1, & \text{for } m \geq 3. \end{cases}$$

One should pay particular notice to the fact that the cumulative distribution function is a monotone nondecreasing function defined not only for the values assumed by the given random variable but for all real numbers.

Example 3.10: Find the cumulative distribution function of the random variable X in Example 3.9. Using $F(x)$, verify that $f(2) = 3/8$.

Solution: Direct calculations of the probability distribution of Example 3.9 give $f(0) = 1/16$, $f(1) = 1/4$, $f(2) = 3/8$, $f(3) = 1/4$, and $f(4) = 1/16$. Therefore,

$$F(0) = f(0) = \frac{1}{16},$$

$$F(1) = f(0) + f(1) = \frac{5}{16},$$

$$F(2) = f(0) + f(1) + f(2) = \frac{11}{16},$$

$$F(3) = f(0) + f(1) + f(2) + f(3) = \frac{15}{16},$$

$$F(4) = f(0) + f(1) + f(2) + f(3) + f(4) = 1.$$

Hence,

$$F(x) = \begin{cases} 0, & \text{for } x < 0, \\ \frac{1}{16}, & \text{for } 0 \leq x < 1, \\ \frac{5}{16}, & \text{for } 1 \leq x < 2, \\ \frac{11}{16}, & \text{for } 2 \leq x < 3, \\ \frac{15}{16}, & \text{for } 3 \leq x < 4, \\ 1 & \text{for } x \geq 4. \end{cases}$$

Now

$$f(2) = F(2) - F(1) = \frac{11}{16} - \frac{5}{16} = \frac{3}{8}.$$

It is often helpful to look at a probability distribution in graphic form. One might plot the points $(x, f(x))$ of Example 3.9 to obtain Figure 3.1. By joining the points to the x axis either with a dashed or with a solid line, we obtain a probability mass function plot. Figure 3.1 makes it easy to see what values of X are most likely to occur, and it also indicates a perfectly symmetric situation in this case.

Instead of plotting the points $(x, f(x))$, we more frequently construct rectangles, as in Figure 3.2. Here the rectangles are constructed so that their bases of equal width are centered at each value x and their heights are equal to the corresponding probabilities given by $f(x)$. The bases are constructed so as to leave no space between the rectangles. Figure 3.2 is called a **probability histogram**.

Since each base in Figure 3.2 has unit width, $P(X = x)$ is equal to the area of the rectangle centered at x. Even if the bases were not of unit width, we could adjust the heights of the rectangles to give areas that would still equal the probabilities of X assuming any of its values x. This concept of using areas to represent
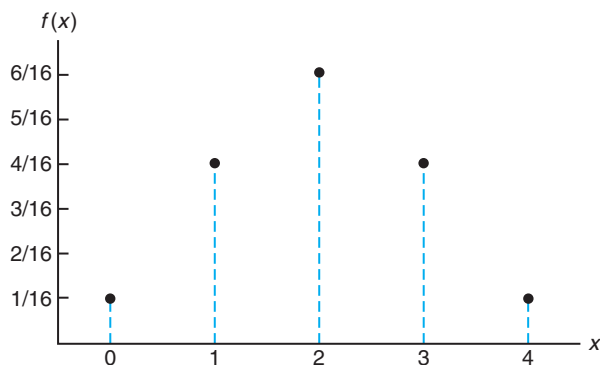
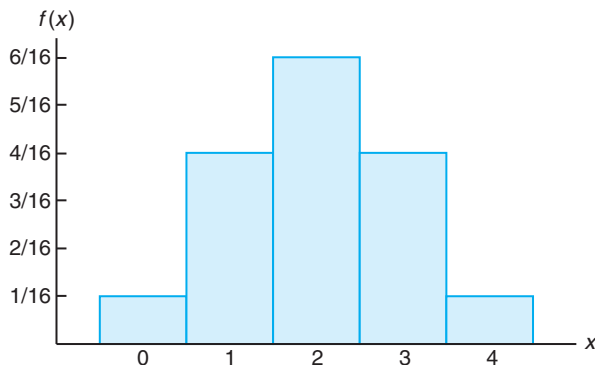Figure 3.1: Probability mass function plot.

Figure 3.2: Probability histogram.

probabilities is necessary for our consideration of the probability distribution of a continuous random variable.

The graph of the cumulative distribution function of Example 3.9, which appears as a step function in Figure 3.3, is obtained by plotting the points $(x, F(x))$.

Certain probability distributions are applicable to more than one physical situation. The probability distribution of Example 3.9, for example, also applies to the random variable Y, where Y is the number of heads when a coin is tossed 4 times, or to the random variable W, where W is the number of red cards that occur when 4 cards are drawn at random from a deck in succession with each card replaced and the deck shuffled before the next drawing. Special discrete distributions that can be applied to many different experimental situations will be considered in Chapter 5.
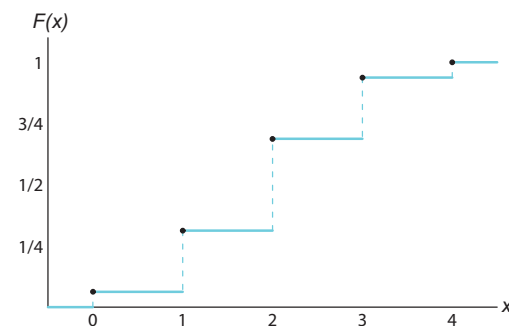

Figure 3.3: Discrete cumulative distribution function.

3.3 Continuous Probability Distributions

A continuous random variable has a probability of 0 of assuming *exactly* any of its values. Consequently, its probability distribution cannot be given in tabular form.

At first this may seem startling, but it becomes more plausible when we consider a particular example. Let us discuss a random variable whose values are the heights of all people over 21 years of age. Between any two values, say 163.5 and 164.5 centimeters, or even 163.99 and 164.01 centimeters, there are an infinite number of heights, one of which is 164 centimeters. The probability of selecting a person at random who is exactly 164 centimeters tall and not one of the infinitely large set of heights so close to 164 centimeters that you cannot humanly measure the difference is remote, and thus we assign a probability of 0 to the event. This is not the case, however, if we talk about the probability of selecting a person who is at least 163 centimeters but not more than 165 centimeters tall. Now we are dealing with an interval rather than a point value of our random variable.

We shall concern ourselves with computing probabilities for various intervals of continuous random variables such as $P(a < X < b)$, $P(W \geq c)$, and so forth. Note that when X is continuous,

$$P(a < X \leq b) = P(a < X < b) + P(X = b) = P(a < X < b).$$

That is, it does not matter whether we include an endpoint of the interval or not. This is not true, though, when X is discrete.

Although the probability distribution of a continuous random variable cannot be presented in tabular form, it can be stated as a formula. Such a formula would necessarily be a function of the numerical values of the continuous random variable X and as such will be represented by the functional notation $f(x)$. In dealing with continuous variables, $f(x)$ is usually called the **probability density function**, or simply the **density function**, of X. Since X is defined over a continuous sample space, it is possible for $f(x)$ to have a finite number of discontinuities. However, most density functions that have practical applications in the analysis of statistical data are continuous and their graphs may take any of several forms, some of which are shown in Figure 3.4. Because areas will be used to represent probabilities and probabilities are positive numerical values, the density function must lie entirely above the x axis.
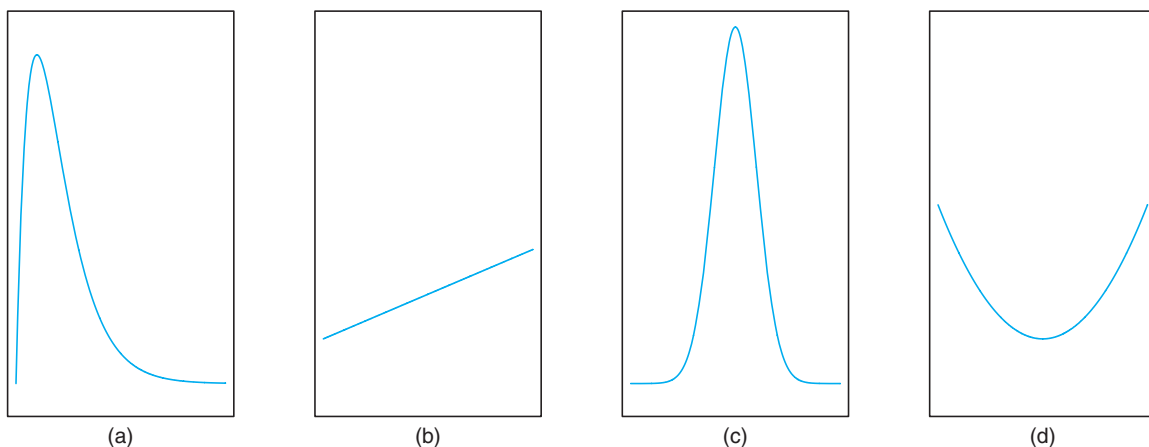

Figure 3.4: Typical density functions.

A probability density function is constructed so that the area under its curve

bounded by the x axis is equal to 1 when computed over the range of X for which $f(x)$ is defined. Should this range of X be a finite interval, it is always possible to extend the interval to include the entire set of real numbers by defining $f(x)$ to be zero at all points in the extended portions of the interval. In Figure 3.5, the probability that X assumes a value between a and b is equal to the shaded area under the density function between the ordinates at $x = a$ and $x = b$, and from integral calculus is given by
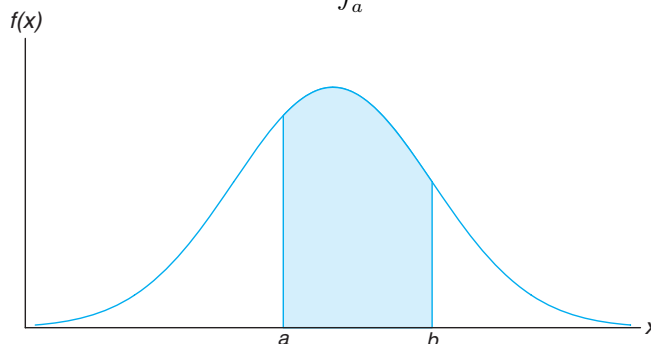
$$P(a < X < b) = \int_a^b f(x) \ dx.$$

Figure 3.5: $P(a < X < b)$.

Definition 3.6: The function $f(x)$ is a **probability density function** (pdf) for the continuous random variable X, defined over the set of real numbers, if

1. $f(x) \geq 0$, for all $x \in R$.

2. $\int_{-\infty}^{\infty} f(x) \ dx = 1$.

3. $P(a < X < b) = \int_a^b f(x) \ dx$.

Example 3.11: Suppose that the error in the reaction temperature, in °C, for a controlled laboratory experiment is a continuous random variable X having the probability density function

$$f(x) = \begin{cases} \frac{x^2}{3}, & -1 < x < 2, \\ 0, & \text{elsewhere.} \end{cases}$$

.

(a) Verify that $f(x)$ is a density function.

(b) Find $P(0 < X \leq 1)$.

Solution: We use Definition 3.6.

(a) Obviously, $f(x) \geq 0$. To verify condition 2 in Definition 3.6, we have

$$\int_{-\infty}^{\infty} f(x) \ dx = \int_{-1}^{2} \frac{x^2}{3} dx = \frac{x^3}{9}\Big|_{-1}^{2} = \frac{8}{9} + \frac{1}{9} = 1.$$

(b) Using formula 3 in Definition 3.6, we obtain

$$P(0 < X \le 1) = \int_0^1 \frac{x^2}{3}\,dx = \frac{x^3}{9}\bigg|_0^1 = \frac{1}{9}.$$

Definition 3.7: The **cumulative distribution function** $F(x)$ of a continuous random variable X with density function $f(x)$ is

$$F(x) = P(X \le x) = \int_{-\infty}^x f(t)\,dt, \quad \text{for } -\infty < x < \infty.$$

As an immediate consequence of Definition 3.7, one can write the two results

$$P(a < X < b) = F(b) - F(a) \text{ and } f(x) = \frac{dF(x)}{dx},$$

if the derivative exists.

Example 3.12: For the density function of Example 3.11, find $F(x)$, and use it to evaluate $P(0 < X \le 1)$.

Solution: For $-1 < x < 2$,

$$F(x) = \int_{-\infty}^x f(t)\,dt = \int_{-1}^x \frac{t^2}{3}\,dt = \frac{t^3}{9}\bigg|_{-1}^x = \frac{x^3 + 1}{9}.$$

Therefore,

$$F(x) = \begin{cases} 0, & x < -1, \\ \frac{x^3+1}{9}, & -1 \le x < 2, \\ 1, & x \ge 2. \end{cases}$$

The cumulative distribution function $F(x)$ is expressed in Figure 3.6. Now

$$P(0 < X \le 1) = F(1) - F(0) = \frac{2}{9} - \frac{1}{9} = \frac{1}{9},$$

which agrees with the result obtained by using the density function in Example 3.11.

Example 3.13: The Department of Energy (DOE) puts projects out on bid and generally estimates what a reasonable bid should be. Call the estimate b. The DOE has determined that the density function of the winning (low) bid is

$$f(y) = \begin{cases} \frac{5}{8b}, & \frac{2}{5}b \le y \le 2b, \\ 0, & \text{elsewhere.} \end{cases}$$

Find $F(y)$ and use it to determine the probability that the winning bid is less than the DOE's preliminary estimate b.

Solution: For $2b/5 \le y \le 2b$,

$$F(y) = \int_{2b/5}^y \frac{5}{8b}\,dy = \frac{5t}{8b}\bigg|_{2b/5}^y = \frac{5y}{8b} - \frac{1}{4}.$$
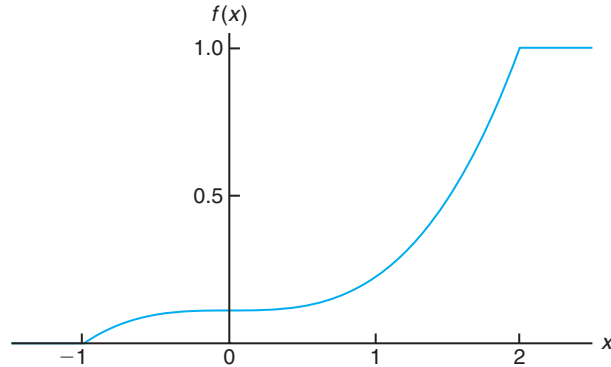
Figure 3.6: Continuous cumulative distribution function.

Thus,

$$F(y) = \begin{cases} 0, & y < \frac{2}{5}b, \\ \frac{5y}{8b} - \frac{1}{4}, & \frac{2}{5}b \le y < 2b, \\ 1, & y \ge 2b. \end{cases}$$

To determine the probability that the winning bid is less than the preliminary bid estimate b, we have

$$P(Y \le b) = F(b) = \frac{5}{8} - \frac{1}{4} = \frac{3}{8}.$$

Exercises

3.1 Classify the following random variables as discrete or continuous:

X: the number of automobile accidents per year in Virginia.

Y: the length of time to play 18 holes of golf.

M: the amount of milk produced yearly by a particular cow.

N: the number of eggs laid each month by a hen.

P: the number of building permits issued each month in a certain city.

Q: the weight of grain produced per acre.

3.2 An overseas shipment of 5 foreign automobiles contains 2 that have slight paint blemishes. If an agency receives 3 of these automobiles at random, list the elements of the sample space S, using the letters B and N for blemished and nonblemished, respectively; then to each sample point assign a value x of the random variable X representing the number of automobiles with paint blemishes purchased by the agency.

3.3 Let W be a random variable giving the number of heads minus the number of tails in three tosses of a coin. List the elements of the sample space S for the three tosses of the coin and to each sample point assign a value w of W.

3.4 A coin is flipped until 3 heads in succession occur. List only those elements of the sample space that require 6 or less tosses. Is this a discrete sample space? Explain.

3.5 Determine the value c so that each of the following functions can serve as a probability distribution of the discrete random variable X:

(a) $f(x) = c(x^2 + 4)$, for $x = 0, 1, 2, 3$;

(b) $f(x) = c\binom{2}{x}\binom{3}{3-x}$, for $x = 0, 1, 2$.

3.6 The shelf life, in days, for bottles of a certain prescribed medicine is a random variable having the density function

$$f(x) = \begin{cases} \frac{20,000}{(x+100)^3}, & x > 0, \\ 0, & \text{elsewhere.} \end{cases}$$

Find the probability that a bottle of this medicine will have a shell life of

(a) at least 200 days;

(b) anywhere from 80 to 120 days.

3.7 The total number of hours, measured in units of 100 hours, that a family runs a vacuum cleaner over a period of one year is a continuous random variable X that has the density function

$$f(x) = \begin{cases} x, & 0 < x < 1, \\ 2 - x, & 1 \le x < 2, \\ 0, & \text{elsewhere.} \end{cases}$$

Find the probability that over a period of one year, a family runs their vacuum cleaner

(a) less than 120 hours;

(b) between 50 and 100 hours.

3.8 Find the probability distribution of the random variable W in Exercise 3.3, assuming that the coin is biased so that a head is twice as likely to occur as a tail.

3.9 The proportion of people who respond to a certain mail-order solicitation is a continuous random variable X that has the density function

$$f(x) = \begin{cases} \frac{2(x+2)}{5}, & 0 < x < 1, \\ 0, & \text{elsewhere.} \end{cases}$$

(a) Show that $P(0 < X < 1) = 1$.

(b) Find the probability that more than 1/4 but fewer than 1/2 of the people contacted will respond to this type of solicitation.

3.10 Find a formula for the probability distribution of the random variable X representing the outcome when a single die is rolled once.

3.11 A shipment of 7 television sets contains 2 defective sets. A hotel makes a random purchase of 3 of the sets. If x is the number of defective sets purchased by the hotel, find the probability distribution of X. Express the results graphically as a probability histogram.

3.12 An investment firm offers its customers municipal bonds that mature after varying numbers of years. Given that the cumulative distribution function of T, the number of years to maturity for a randomly selected bond, is

$$F(t) = \begin{cases} 0, & t < 1, \\ \frac{1}{4}, & 1 \le t < 3, \\ \frac{1}{2}, & 3 \le t < 5, \\ \frac{3}{4}, & 5 \le t < 7, \\ 1, & t \ge 7, \end{cases}$$

find

(a) $P(T = 5)$;

(b) $P(T > 3)$;

(c) $P(1.4 < T < 6)$;

(d) $P(T \le 5 \mid T \ge 2)$.

3.13 The probability distribution of X, the number of imperfections per 10 meters of a synthetic fabric in continuous rolls of uniform width, is given by

x	0	1	2	3	4
$f(x)$	0.41	0.37	0.16	0.05	0.01

Construct the cumulative distribution function of X.

3.14 The waiting time, in hours, between successive speeders spotted by a radar unit is a continuous random variable with cumulative distribution function

$$F(x) = \begin{cases} 0, & x < 0, \\ 1 - e^{-8x}, & x \ge 0. \end{cases}$$

Find the probability of waiting less than 12 minutes between successive speeders

(a) using the cumulative distribution function of X;

(b) using the probability density function of X.

3.15 Find the cumulative distribution function of the random variable X representing the number of defectives in Exercise 3.11. Then using $F(x)$, find

(a) $P(X = 1)$;

(b) $P(0 < X \le 2)$.

3.16 Construct a graph of the cumulative distribution function of Exercise 3.15.

3.17 A continuous random variable X that can assume values between $x = 1$ and $x = 3$ has a density function given by $f(x) = 1/2$.

(a) Show that the area under the curve is equal to 1.

(b) Find $P(2 < X < 2.5)$.

(c) Find $P(X \le 1.6)$.

3.18 A continuous random variable X that can assume values between $x = 2$ and $x = 5$ has a density function given by $f(x) = 2(1 + x)/27$. Find

(a) $P(X < 4)$;

(b) $P(3 \leq X < 4)$.

3.19 For the density function of Exercise 3.17, find $F(x)$. Use it to evaluate $P(2 < X < 2.5)$.

3.20 For the density function of Exercise 3.18, find $F(x)$, and use it to evaluate $P(3 \leq X < 4)$.

3.21 Consider the density function

$$f(x) = \begin{cases} k\sqrt{x}, & 0 < x < 1, \\ 0, & \text{elsewhere.} \end{cases}$$

(a) Evaluate k.

(b) Find $F(x)$ and use it to evaluate

$$P(0.3 < X < 0.6).$$

3.22 Three cards are drawn in succession from a deck without replacement. Find the probability distribution for the number of spades.

3.23 Find the cumulative distribution function of the random variable W in Exercise 3.8. Using $F(w)$, find

(a) $P(W > 0)$;

(b) $P(-1 \leq W < 3)$.

3.24 Find the probability distribution for the number of jazz CDs when 4 CDs are selected at random from a collection consisting of 5 jazz CDs, 2 classical CDs, and 3 rock CDs. Express your results by means of a formula.

3.25 From a box containing 4 dimes and 2 nickels, 3 coins are selected at random without replacement. Find the probability distribution for the total T of the 3 coins. Express the probability distribution graphically as a probability histogram.

3.26 From a box containing 4 black balls and 2 green balls, 3 balls are drawn in succession, each ball being replaced in the box before the next draw is made. Find the probability distribution for the number of green balls.

3.27 The time to failure in hours of an important piece of electronic equipment used in a manufactured DVD player has the density function

$$f(x) = \begin{cases} \frac{1}{2000} \exp(-x/2000), & x \geq 0, \\ 0, & x < 0. \end{cases}$$

(a) Find $F(x)$.

(b) Determine the probability that the component (and thus the DVD player) lasts more than 1000 hours before the component needs to be replaced.

(c) Determine the probability that the component fails before 2000 hours.

3.28 A cereal manufacturer is aware that the weight of the product in the box varies slightly from box to box. In fact, considerable historical data have allowed the determination of the density function that describes the probability structure for the weight (in ounces). Letting X be the random variable weight, in ounces, the density function can be described as

$$f(x) = \begin{cases} \frac{2}{5}, & 23.75 \leq x \leq 26.25, \\ 0, & \text{elsewhere.} \end{cases}$$

(a) Verify that this is a valid density function.

(b) Determine the probability that the weight is smaller than 24 ounces.

(c) The company desires that the weight exceeding 26 ounces be an extremely rare occurrence. What is the probability that this rare occurrence does actually occur?

3.29 An important factor in solid missile fuel is the particle size distribution. Significant problems occur if the particle sizes are too large. From production data in the past, it has been determined that the particle size (in micrometers) distribution is characterized by

$$f(x) = \begin{cases} 3x^{-4}, & x > 1, \\ 0, & \text{elsewhere.} \end{cases}$$

(a) Verify that this is a valid density function.

(b) Evaluate $F(x)$.

(c) What is the probability that a random particle from the manufactured fuel exceeds 4 micrometers?

3.30 Measurements of scientific systems are always subject to variation, some more than others. There are many structures for measurement error, and statisticians spend a great deal of time modeling these errors. Suppose the measurement error X of a certain physical quantity is decided by the density function

$$f(x) = \begin{cases} k(3 - x^2), & -1 \leq x \leq 1, \\ 0, & \text{elsewhere.} \end{cases}$$

(a) Determine k that renders $f(x)$ a valid density function.

(b) Find the probability that a random error in measurement is less than $1/2$.

(c) For this particular measurement, it is undesirable if the *magnitude* of the error (i.e., $|x|$) exceeds 0.8. What is the probability that this occurs?

3.31 Based on extensive testing, it is determined by the manufacturer of a washing machine that the time Y (in years) before a major repair is required is characterized by the probability density function

$$f(y) = \begin{cases} \frac{1}{4}e^{-y/4}, & y \geq 0, \\ 0, & \text{elsewhere.} \end{cases}$$

(a) Critics would certainly consider the product a bargain if it is unlikely to require a major repair before the sixth year. Comment on this by determining $P(Y > 6)$.
(b) What is the probability that a major repair occurs in the first year?

3.32 The proportion of the budget for a certain type of industrial company that is allotted to environmental and pollution control is coming under scrutiny. A data collection project determines that the distribution of these proportions is given by

$$f(y) = \begin{cases} 5(1-y)^4, & 0 \leq y \leq 1, \\ 0, & \text{elsewhere.} \end{cases}$$

(a) Verify that the above is a valid density function.
(b) What is the probability that a company chosen at random expends less than 10% of its budget on environmental and pollution controls?
(c) What is the probability that a company selected at random spends more than 50% of its budget on environmental and pollution controls?

3.33 Suppose a certain type of small data processing firm is so specialized that some have difficulty making a profit in their first year of operation. The probability density function that characterizes the proportion Y that make a profit is given by

$$f(y) = \begin{cases} ky^4(1-y)^3, & 0 \leq y \leq 1, \\ 0, & \text{elsewhere.} \end{cases}$$

(a) What is the value of k that renders the above a valid density function?
(b) Find the probability that at most 50% of the firms make a profit in the first year.
(c) Find the probability that at least 80% of the firms make a profit in the first year.

3.34 Magnetron tubes are produced on an automated assembly line. A sampling plan is used periodically to assess quality of the lengths of the tubes. This measurement is subject to uncertainty. It is thought that the probability that a random tube meets length specification is 0.99. A sampling plan is used in which the lengths of 5 random tubes are measured.

(a) Show that the probability function of Y, the number out of 5 that meet length specification, is given by the following discrete probability function:

$$f(y) = \frac{5!}{y!(5-y)!}(0.99)^y(0.01)^{5-y},$$

for $y = 0, 1, 2, 3, 4, 5$.

(b) Suppose random selections are made off the line and 3 are outside specifications. Use $f(y)$ above either to support or to refute the conjecture that the probability is 0.99 that a single tube meets specifications.

3.35 Suppose it is known from large amounts of historical data that X, the number of cars that arrive at a specific intersection during a 20-second time period, is characterized by the following discrete probability function:

$$f(x) = e^{-6}\frac{6^x}{x!}, \text{ for } x = 0, 1, 2, \ldots.$$

(a) Find the probability that in a specific 20-second time period, more than 8 cars arrive at the intersection.
(b) Find the probability that only 2 cars arrive.

3.36 On a laboratory assignment, if the equipment is working, the density function of the observed outcome, X, is

$$f(x) = \begin{cases} 2(1-x), & 0 < x < 1, \\ 0, & \text{otherwise.} \end{cases}$$

(a) Calculate $P(X \leq 1/3)$.
(b) What is the probability that X will exceed 0.5?
(c) Given that $X \geq 0.5$, what is the probability that X will be less than 0.75?